

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Exploring user behavior and needs in Q & A communities

### Conference or Workshop Item

#### How to cite:

Choudhury, Smitashree and Alani, Harith (2014). Exploring user behavior and needs in Q & A communities. In: Proceedings of the European Conference on Social Media: ECSM 2014, Academic Conferences and Publishing International Limited, pp. 80–89.

For guidance on citations see [FAQs](#).

© 2014 The Authors

Version: Accepted Manuscript

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# Exploring User Behavior and Needs in Q & A Communities

Smitashree Choudhury & Dr. Harith Alani

KMI, UK Open University

[Smitashree.choudhury@open.ac.uk](mailto:Smitashree.choudhury@open.ac.uk)

[harith.alani@open.ac.uk](mailto:harith.alani@open.ac.uk)

## Abstract

One of the difficult challenges of any knowledge centric online community is to sustain the momentum of knowledge sharing and knowledge creation effort by its members through various means. This requires a clearer understanding of user needs that drive community members to contribute, engage and stay loyal to the community. In this paper, we explore the applicability of Abraham Maslow's theory (1943) to understand user behavior and their latent needs using Exploratory Factor analysis. Results show that users are largely driven by four main needs: social interaction, altruism cognitive need and reputation. Our results further indicate that users with high reputations are more likely to stay longer in the community than others, and that socially motivated users are responsible for increased content creation.

**Keywords:** Online Communities, User behavior.

## 1. Introduction

Many organisations are now taking serious note of managing their online communities, which are fast becoming knowledge hubs for their employees as well as for customers. Despite the huge success of virtual communities as communication tool, little is known how and why community users participate and contributes. Active participation, quality content creation are crucial for the viability of content based online communities (Koh et.al,2007). Based on this premise, researchers have started identifying various motivations of user participation and contribution in such communities (Nov et al, 2008). Identifying the motivations that drive user participation, engagement and contribution would help community managers, developers, and analysts to gain insights into how these communities thrive and survive. A clear understanding of user motivation will not only help community managers for efficient management, but will also provide great benefits to system designers in developing dynamic and self-adapted online social systems.

In this paper we focus on Question and Answer (Q&A) communities in an enterprise setup where users create, share, discuss issues ranging from product development, services, technical support etc. It allows users to follow other users, award points to other users for their contributions. In particular we address the following research questions: *What different user needs are satisfied from community participation and contribution? how do these different needs correlate with user behavior?*, and finally *do the needs and their evolution, follow structural map of Maslow's hierarchical need theory?*

Hereafter, we begin with a literature study of the area in section two, section three describes the model mapping followed by experimental details in section four. Finally we conclude with few limitations in section five.

## 2. Related Work

Many existing studies have investigated the motivations for online participation and contribution suggesting a wide range of personal and social factors (fun, knowledge seeking, social identity, esteem etc.) as reasons for online participation and contribution. Existing literature in this area can be broadly organized in two categories; (1) investigations on the use of social theories to understand user motivation, and (2) research on method of study, e.g. survey and questionnaire vs. data centric analysis methods. We will briefly describe example studies from these two categories.

### **2.1 Use of social theories to understand online user motivation.**

User behavior and motivation to participate and contribute in online communities has been grounded with various existing social theories ranging from Uses and Gratification theory (U&G), collective

action theory, self-determination theory, theory of reciprocity, social identity theory to name a few. Lampe et.al (2010) used U&G theory to explain the influence of belongingness, social and cognitive factors in user participation, while Dholakia and colleagues (2004) studied the motivational role of group norms and social identity and suggested six benefits for users including information seeking, sharing and reputation. Wask and Faraz (2005) used the theory of collective action to explain the motivational influence of expected reputation in contributing to professional forums. Studies by (Hars & Ou, 2002) described the role of intrinsic and extrinsic motivators in content contribution. Krasnova et. al.(2008) suggested need for belongingness, esteem and peer pressure as prime motivators for participation and contribution. Burke and Lento (2009) emphasized the positive role of social learning and feedback on new users in Facebook. Chiu and colleagues (2006) used Social Cognitive Theory and Social Capital Theory to explain the impact of social ties, reciprocity and identity on users' contribution. Joinson (2008) identified seven uses and gratification (U&G) of Facebook use. These themes are social connection, shared identities, content, social investigation, social network surfing and status updating. Contribution to open source projects are motivated by self-development, reputation and altruism (Oreg et. al., 2008) while fun and ideology proved to be the prime motivators for Wikipedia contribution (Nov, 2007).

## **2.2 Self-report vs. Data driven approach**

Majority of studies on user motivation follow the self-reported feedback method to collect data about a user's reason to participate as compared to data driven approaches where server logs are used to analyse user behavior to infer motivation. Self-reported approaches carry the limitations of sample size, recall bias (Brewer, 2000), and bias of social desirability (Crowne & Marlowe, 1960). Lately, data driven approach attracted researchers attention due to easy availability of large amount of user data. A combination of user interviews and server log data of Knowledge-iN; a large South Korean Q&A community, was used in the study by Nam et.al. (2009). This study revealed five motivations for contribution, including helping others, self-promotion, learning, recreation and reputation points. Similar to this study, in this paper we use a data-centric approach, where users' actual activities and interactions in the community are extracted, and used for a pragmatic correlation between user needs and behaviour.

## **3. Mapping Maslow's Hierarchy to Online Communities**

Maslow's theory on human needs and motivation provides a powerful theory of human behavior. He proposed five different needs that drive human behavior at every stage of life depending on the satisfaction of the most pre-potent needs. These needs often visualised as a pyramid to reflect their order and satisfiability quotient. In this section we describe how we map Maslow's pyramid to the domain of online communities (figure 1). This mapping enables us to study the needs of users in online communities in light of Maslow's needs hierarchy.

*Physiological and Security needs:* are at the bottom of the hierarchy and are considered very basic needs for survival, which includes the need for food, housing, etc. In the context of the online world, these needs may be translated into system access, hardware requirements such as computer, community access, online identity, etc. We presume that these needs are already met when users join an online community, and hence they not the focus of this paper.

*Need for Belongingness:* reflects users' desire to be part of the community, have interpersonal relationship and a sense of acceptance from their social group. In the context of social media, this need may be translated into a need for connection, making friendships, being part of an interest group etc. In Q&A communities, connection is made through replies, comments and voting, which can be considered as a proxy for the desire to establish such social connections.

*Need for self-esteem:* According to Maslow, the need for reputation and self-esteem emerges once the individual is settled with his social identity through groups and communities. For online communities, reputation seems to be one of the strong motivators in many previous studies (see section 2) It makes intuitive sense that users of professional and Q&A communities would wish to be recognized among their peer groups, and hence their desire to excel could be reflected by specific community behavior, such as answering more questions, attempting complex questions etc..

*Need for Self-actualisation:* Maslow's original theory proposed self-actualisation as a difficult phase to reach. This stage is characterised by attributes such as efficient perception of reality, creativity,

spontaneity in ideas and actions, interest in helping others (altruism), etc. Although it is difficult to claim that online communities satisfy user's self-actualisation need, such social environments enable more users to be helpful to others, e.g., by replying to other's questions, and contributing towards the community's benefits. Here we focus on the characteristic of altruism.

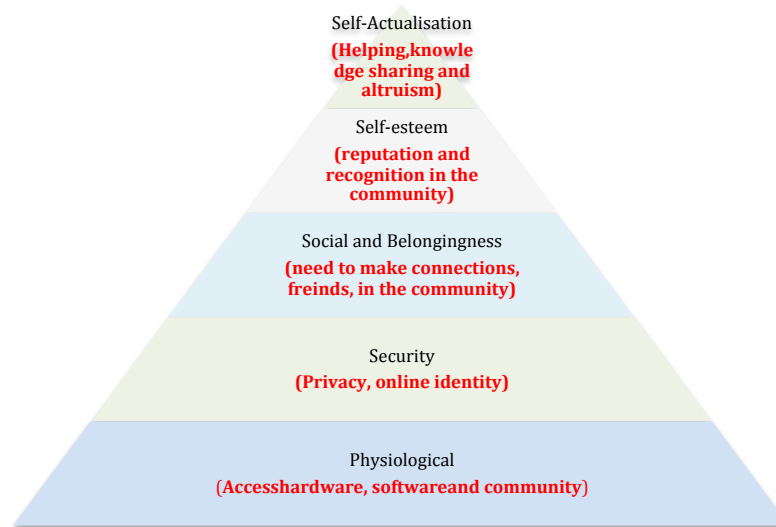


Figure 1: Maslow's Pyramid and our mapping (in red) to online Q&A communities.

This study involves two subsections (1) Factor analysis of user features in order to identify possible need factors and (2) analyse the evolution of need factors over time.

## 4. Experiment

### 4.1 Dataset and Feature Engineering

To ground our work, we used SAP community network (SCN) for user behavior analysis and need identification. SAP community network is a collection of forums focusing on various SAP related products, services hosted by SAP. SCN has a reputation system where users are awarded points and badges for their quality contribution. The snapshot of data provided for this work consists 34 different forums with 95200 threads and 427000 posts from 32926 users.

Need and behavior are often confused and used interchangeably. A finer distinction exist between these two concepts where need is considered subjective and non-observable while behavior is observable and taken as external manifestation of internal need. To measure needs we need to measure behavioral intensities, accordingly we extracted features relevant to users within an online community:

- *Community Age* is the duration of time user is active in the community.
- *Forum Focus* indicates dispersion of users attention between number of forums within the community. A higher score indicates wide focus while a small score indicates concentration..
- *Post frequency (PPM)*: number of posts created by a user per time interval (here in a month).
- *Initiation share*: proportion of threads started by a user in the community.
- *Reply share*: proportion of replies given by a user in the community.
- *Initiation ratio*: user's ratio of initiation to his replies.
- *Reply ratio*: user's ratio of replies to his initiation
- *Self-reply ratio*: user's ratio of replies directed towards ones own initiated thread.
- *Normalized Content Quality (NCQ)*: indicates the average score a user gets for each contribution (total number of points / number of posts). We use NCQ as a reflection of user *reputation* .
- *In-degree*: proportion of unique users replied to user, alternatively termed here as "popularity".
- *Out-degree*: proportion of unique users that user has replied to, alternatively termed here as "engagement".
- *Between-ness centrality*: degree of centrality of a user within the reply network.

- *Tie strength*: indicates the strength of interactions of a user ranging between 0-1.
- *Topic Focus*: High score indicates, spread while low score is an indication of focus.

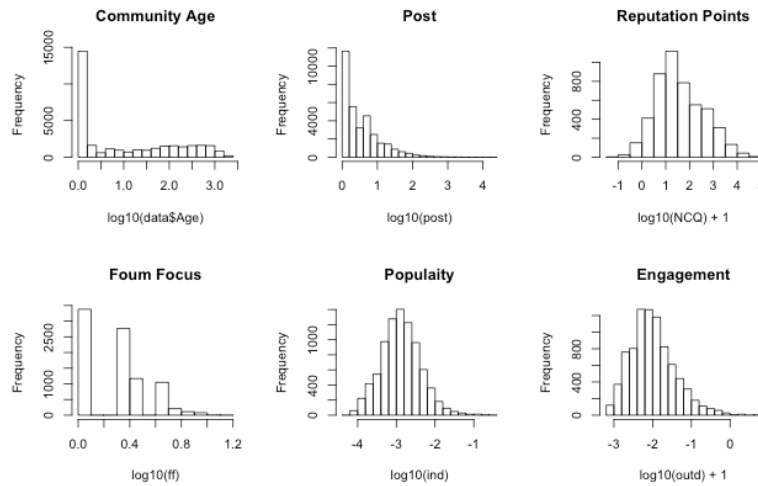


Figure 2: Distribution of some major user related features observed in SAP dataset.

To clarify further, we plot the distribution of important attributes such as community age, number of posts per user, reputation points, forum focus, popularity and engagement scores of community users in Figure 2. Despite many other variations, most of the behavioural features are characterised by a common pattern of heavy tailed distribution; further indicating dominance of specific features for certain cluster of users.

## 4.2 Features To Factors

Mapping user features to any motive or need is non-trivial. Each motive/need may be reflected through one or more user features. In order to get a better understanding of these features and how they correlate with each other we, use the Exploratory factor Analysis approach. Exploratory Factor Analysis (EFA) is a multivariate statistical approach used in social science research for factor discovery by reducing a large number of variables into a smaller set of variables (factors). EFA involves five fundamental steps;

- (1) Feature Correlation.
- (2) Number of factors to be extracted.
- (3) Method to extract factors.
- (4) Choosing a rotation method.
- (5) Interpretation and factor labeling

**4.2.1 Correlation Matrix:** Inter-feature correlation (figure 1) shows that features exhibit both negative and positive relationship with different degrees while some features seem to be independent.

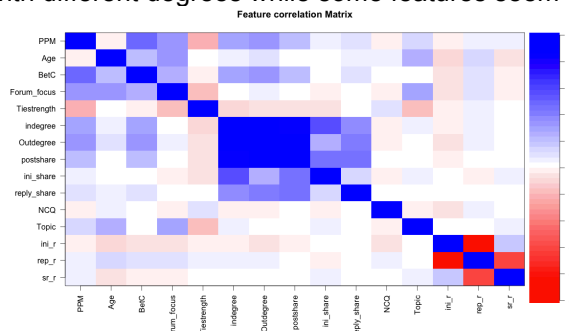
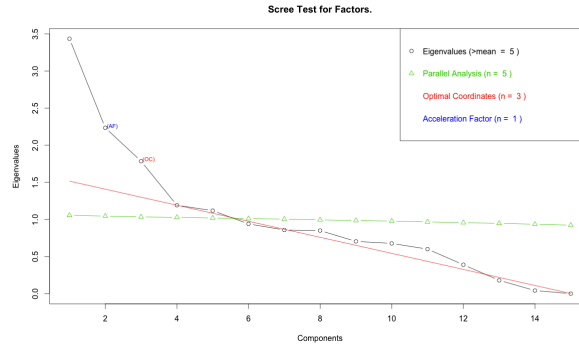


Figure 3: Correlation matrix between user features.

The correlation matrix (Fig. 3) reveals a weak positive relationship between a user's reputation and reply behaviour ( $r = .17$ ) while it is nearly un-correlated ( $r = .01$ ) with overall contribution volume.

However, this unexpected lack of correlation is not statistically significant. The strongest correlation for contribution volume is observed with social attributes such as in-degree, out-degree distribution and centrality measure ( $r = .25$ ,  $.29$  and  $.82$  respectively,  $p = 0.001$ ). High topic entropy is also positively related to [the overall contribution](#) ( $r = .55$ ,  $p = 0.01$ ).

**4.2.2 Number of factors to be extracted:** Decision on how many factors need to be extracted is mostly subjective and explained by multiple criteria e.g. scree test (Cattell 1966), parallel analysis (Thompson 1996), Kaiser's (Kaiser, 1960) Eigen value criteria ( $>1.0$ ) and theoretical perspective. Given the available choices and their nuance differences, Thompson and Daniel (1996) suggested simultaneous use of multiple criteria for an ideal solution. We opted for scree test (fig. 4), Eigen values threshold and parallel analysis to select the number suggested by the majority of the approaches, in our case it is 4 to 5.



**Figure 4:** A scree plot showing number of factors to be extracted from the list of features.

**4.2.3 Factor extraction method:** Maximum likelihood, Principal axis factoring (PAF), and Principal Component analysis (PCA) are some of the known factor extraction techniques. Each method aims to reduce the number of observed variables into groups of correlated variables. The most popular methods are PAF and PCA (Henson, 2006). Although both techniques give mostly identical results in terms of factor discoverability, their underlying mechanism to group variables differ. While PCA takes into account both unique and shared variances between observed variables, PAF only considers the shared variances. We decided to use both approaches in our experiment to get a broader picture.

**4.2.4 Rotation Method:** Individual features may be loaded onto more than one factor making the result difficult to interpret. Hence Factor analysis involves rotation techniques to maximize the high loading items and minimize the low loading variables and making factor interpretation more reliable. There are two categories of rotation techniques; (1) orthogonal and (2) oblique rotation. Orthogonal rotation produces uncorrelated factor structure while oblique rotation treats factor as correlated. From each method we have multiple options (varimax, quartimax, oblimin etc.) to choose from depending on the data requirements.

Regardless of any rotation and factor extraction method, the objective is to produce a more interpretable and conceptually suitable solution. As per the suggestion of Pett, Lackey, and Sullivan (2003) we tested both the rotation and factor extraction techniques to find the best fit.

**4.2.5 Interpretation and factor labeling:** Following the multiple criteria suggestion to determine the number of factors to be extracted, we decided to extract 5 factors (table 1). Next, we run the factor analysis to get the loadings for each of the factors. PAF using oblique rotation explained 54% of variance through four factors while analysis using PCA resulted in 65.1% (.651) of data variance explained. In the analysis using PAF user reputation (NCQ) did not load on any factors. Absence of reputation related factor is surprising since reputation seems to be a strong motivator, especially in professional communities (Lakhani et al, 2005). However, with the second analysis using PCA and varimax rotation, reputation is loaded onto a single factor accounting for 7% of variance. We selected those features with a loading threshold of  $>.4$  and features that had cross-loaded significantly were discarded. These five factors represent different aspect of user behavior in the community.

**Factor 1** pre-dominantly includes features concerning user's social network properties such as in-degree, out-degree. It is not surprising to see that users high on this factor are high in their overall contribution (proportion of individual posts in relation to the community posts). We label this factor as one that belongs to "Socially active users/engagers".

**Factor 2** comprised of three features related to the user's contribution behavior e.g. initiation, reply and self-reply. We found that both PAF and PCA consider this as the second most important factor with 18% and 14% variance respectively but the direction differs. In factor analysis with PAF this factor gets a positive loading from the reply ratio and negative loadings from both initiation and self-reply while the directions are exactly opposite in PCA. Nonetheless in both cases the factor is clearly focused on contributing behavior of the user. These features reflect the purpose of contribution whether the contributions are of type information seeking or information or knowledge sharing thereby helping others. We term this factor as "*Askers and Repliers*".

**Factor 3** related to the user's activity frequency such as number of posts per month (PPM), number of forums he participates (forum focus) and his overall position (between-ness centrality) in the community network. It also loads "community age" negatively (-.19) and some degree of "out-degree" but below the threshold. This factor most probably indicates those short term users who come to the community for specific purpose and wanted to put forward their point as much as possible by frequent posting and multiple forum visits. We label this factor explaining "*Active users*".

**Factor 4** contains feature related to experience (high loading of community age=.729), high topic spread and forum focus. We label this factor as "*Experienced users*".

**Factor 5** loads with tie strength and reputation score (NCQ=.618). Reputation score could not be loaded with PAF analysis may be because of its lack of linear correlation with other features except Age (figure 1), but its unique variance is captured by PCA and loaded as the 5<sup>th</sup> factor along with a moderate loading of community age (.339) and suggests a need for recognition and appreciation. We term this factor as "*Reputation/expert users*".

**Table 1:** Factor loadings using PCA

Principal Components Analysis					
PC1	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
In-degree	.924				
Out-degree	.864				
Post share	.901				
Initiation ratio		0.813			
Reply ratio		-0.994			
Self reply ratio		0.686			
PPM			0.800		
BetC			0.791		
Age				0.729	
Forum focus			0.420	0.630	
Topic				0.714	
Tie strength					0.614
NCQ					0.618

As an evaluative measure we computed Chronbach alpha (measures reliability and internal consistency of features as constituent elements of factors) of features, which ranges from .61 to .69 with an average of .63. The findings support the existence of motivation for interaction (factor 1), reputation (factor 5), helping (factor 2) and information seeking (factor 2), while factor 3 and 4 reflects user's activity pattern along with the experience dimension. The next logical goal is to study how these factors evolve over time.

### 4.3 Need Evolution

We recomputed the factor score as following for the evolution study:

- For factor 1 we took the mean score of "out-degree" and in-degree" as the engagement score (ENG).
- NCQ for reputation factor score
- Combined "initiation ratio" and "self-reply ratio" as the information need score (IN).
- Used "Reply ratio" as the helping need/altruism score /community contribution (CC).

We first examined the macro (community) level need evolution to understand what kind of needs are expressed collectively by users in different point of time and their intensity.

The first step in the temporal analysis is the construction of time segments covering the relevant time span (the time for which the data is available). To do this we divided the time period into equal time intervals (16 week each) starting from 2004 to 2010. The start of the first period ( $t_i$ ) would be the beginning of January 2004 to end of April 2004 ( $t_i+16$ ) and the second time period is from  $t_i+16$  to  $t_i+32$ . Overall, this led to 22 time intervals. Each time segment contains the normalized factor score for each user for four factors.

#### 4.3.1 Need Pattern Extraction:

User needs are neither exclusive nor explicit, they appear in combination with other needs with varying degrees for e.g, users with high social interaction may also have a high score on community contribution. This motivated us to extract typical need patterns observed during the time interval  $t_i$ . Our approach to extract need patterns of a time interval considers the relative contribution of individual factor during the time period  $t_i$  – e.g. high information need, low helping need, low engagement.

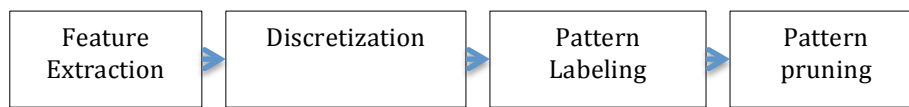


Figure 5: Process of extracting needs patterns from user features.

Figure 5 shows an overview of how we extracted time based need patterns from the user features described above which led us to represent a user with a  $22 \times 4$  feature vector. Next we took the feature score and discretized them by dividing the range into three intervals (1-3) of “high”, “medium” and “low” levels. We also added two more levels (0,4) to represent 0% and 100% because of the nature of the feature computed (features reflecting ratios). The next step is to assign need pattern labels for each time interval corresponding to the feature levels:

IN=Low, CC=High, EN=Medium and Rep=Low -> Need pattern Label (1321)  
 IN=0, CC=1.0, EN=Medium and Rep=Low -> Need pattern Label (0421)

The last stage is the stage of *pattern pruning and categorical labeling*, where we investigated the pattern frequency. A simple frequency count led to 40 unique patterns (figure 6) with 20% of patterns covering 83% of the total distribution.

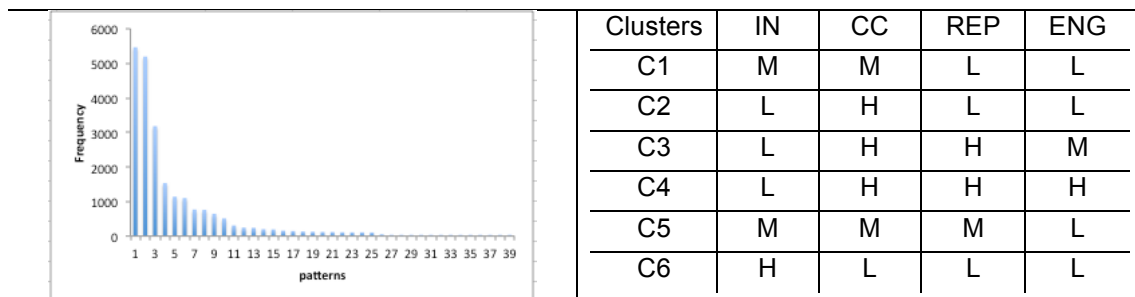


Figure 6: Frequency distribution of unique need patterns.

Table 2: shows 6 clusters with different levels of factor scores.

Initially derived 40 labels is a large number for any meaningful pattern analysis and will result in over-fitting the data hence we further moved to cluster these patterns in order to get a smaller subset by mean of k-means clustering. Clustering of data requires to estimate the number of clusters( $k$ ). We used average silhouette to estimate the number of clusters incrementally starting from 3 to 10 and recording the silhouette coefficient. We took the average silhouette of all the items and compared with different numbers of  $K=\{3...10\}$ , final result showed  $K=6$  with an average silhouette of .53. With 6 clusters, we created categorical label for each cluster depending on feature dimension:

1. **Information seeking and sharing (IR):** user with balanced initiation and reply behavior, low in reputation and low-to medium in engagement.



2. **Information sharing (REP):** Low information, high helping-low reputation, low engagement.
3. **Information Sharing and gain reputation (RR):** users with high reply behavior (100%), no initiation behavior, and medium to high reputation and medium to high engagement.
4. **Information sharing, gain reputation and community engagement (RRE):** High contribution towards the community, high reputation and medium to high engagement.
5. **Information seeking, sharing and gain reputation (IRE):** users with medium initiation and contribution, medium to high reputation and low engagement.
6. **Information Seeking (IN)** users with this label are high on initiation, low on contribution towards other users, low engagement and low reputation.

The cluster output suggests the order of dominance is of 1,6,3,2,4 and cluster 5 and the need patterns of first 3 clusters (1,6, and 3) take approximately 70% of users time.

#### 4.3.2 Need Evolution at the Community Level

Communities' activity trajectory reflects the collective needs of its user base. Figure 7 shows the community level need evolution over the time periods. We computed a cross entropy for each time interval to measure the fluctuation between different times with the following:

$$H(p, q) = - \sum_x p(x) \log q(x)$$

Cross entropy of each time interval shows the amount of fluctuation experienced by the community as a whole decreases with time leading towards a convergence. Indicating the importance of all different needs irrespective of user numbers and activity volume, thus stressing the requirement to examine the need trajectory at the user level and its evolution from initial to final stage.

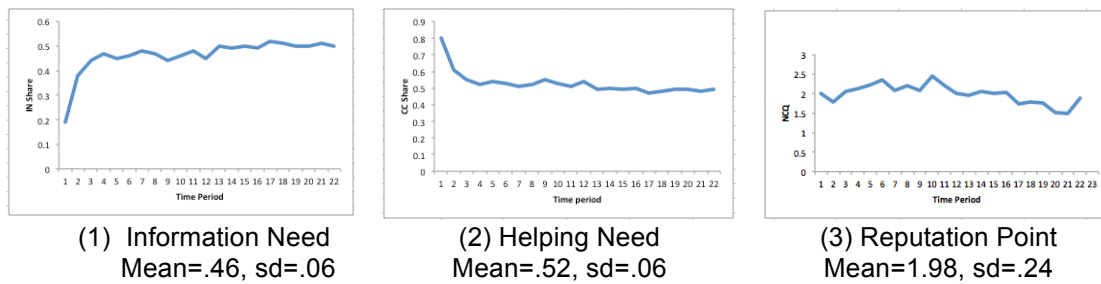


Figure 7: Community level changes in different factor scores.

#### 4.3.3 Need Evolution at the User Level

After joining the community, a user will attempt to address the reason for which he /she joined the community, for example, a user motivated to learn a new skill will start creating content within the forum by posting a questions, requesting for help. while if motivated by knowledge sharing, he will initiate his activities by commenting to the unresolved questions. Continuation in the community depends on satisfaction of his initial motivation, In case of continuation, a user most likely to be engaged in other community related activities requiring more time and effort. Rather than capturing individual user's need evolution, we are interested in the evolution of need patterns. However for the sake of the concreteness, we have illustrated one example of individual user. Figure 8 shows the changes in different need patterns starting from a simple cognitive oriented behavior pattern to community-focused behavior along with high reputation score and wider engagement.

The objective of this evolutionary study is three folds: (1) study the initial need patterns; (2) examine the ending need patterns and (3) the amount of difference observed in-between. Users follow 1 to 6 unique patterns during their life span with different degree of distribution with mean=2 (sd=.74). Distribution of unique need patterns shows that 21% of users are with single need pattern (of which IR takes 71% of the share followed by users with RR 13%) with a mean time interval of 3.36 whereas 51% exhibit 2 unique patterns during their community life span with a mean time interval of 4.36. A positive correlation( $r=.43$ ) between user's number of unique need patterns and the community age suggest community life span increases with multiple need satisfaction.

**Initial Need Patterns:** Ideally, each user should start their online community activities with the lower possible need on the stack, however analysis shows a different picture. Each user exhibits different need patterns when joining the community. Initial need pattern will reflect the initial motivation of a user to join the community. As shown (figure 9) in SAP dataset, 16% of users start with information

seeking need (IN) while 51% to users participate interactively by both initiating and contributing to other users. 5% of users initiate their activity to help others or share their knowledge with a mean time interval of 5.38. Users (12%) starting with RR need patterns stay in the community on average 5.08 intervals but mostly focused on their current status of replying and getting reputation score (mean unique need patterns=1.8).

**Ending Need Pattern:** To investigate how users end their community life and how it differs from their initial stage, we selected those users whose last activity was recorded in 2009; assuming that a complete absence of 1 year from the community indicates either the user left the community or is very infrequent. Users last pattern distribution (Figure 9) shows a similar share with IR being the most frequent followed by IN and RR.

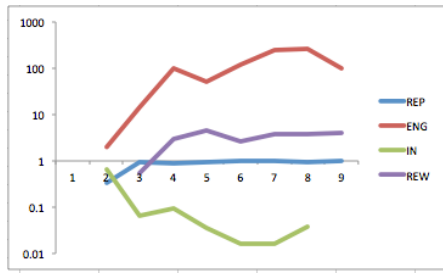


Figure 8: shows the progression of different need factors across 9 time intervals.

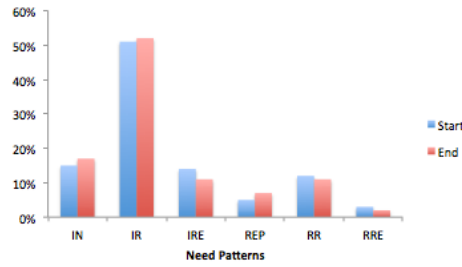


Figure 9: shows the distribution of various need patterns both at the initial (blue) and end stage(red) of user's community life.

**Need Progression:** In order to gain further insight into the progression process from the initial stage to the final stage, we computed a *need progression score* (NPS) for all the intervals of each user. In order to compute NPS, we need to rank the need patterns by means of their distribution frequency. Instead of individual ranking we grouped these patterns into lower and higher order needs based on their average frequency distribution in the data. As a result, IN, IR and IRE are in high frequency group (HFG) while RR, IRE and RRE come under the second group (LFG) because of their low frequency among the users. As in the information retrieval domain, where a high frequency word is considered less relevant while a low frequency word gets higher weight, we ranked the need patterns of high frequency group lower than the need patterns of low-frequency group. Hence a move into the high frequency group will yield a score of -1 while move into the low frequency group will get a score of +1. Following the ordering of need patterns, we computed the average need progression score of user  $nps_m(u_j)$  as follows:

$$NPSm(u_j) = \sum_{i=1}^{N=ai(u_j)} nps_{ti}/N$$

where  $nps(t_i)$  is the need progression score of the interval  $t_i$  computed in relation to the previous interval ( $t_{i-1}$ ),  $ai(u_j)$  is the total number of time intervals user  $j$  has in the community. For each user  $j \in J$ , we computed the absolute need progression score ( $nps_a(u_j)$ ): is the directional difference of last  $nps(t_n)$  from the initial  $nps(t_1)$ .

Analysis of the progression scores shows 46% of users maintain same order of needs during their entire community life while 25% moves from lower to higher order and 28% moves in the reverse direction (high to low). *This finding suggests that users do not follow a rigid hierarchy.*

## 5. Conclusion

We study the applicability of Abraham Maslow's motivation theory to understand the correlation between user behavior and needs in online Q&A communities. Unlike most previous studies, our work is mainly based on behavioural data logged in community systems. Our results suggest that online communities serve several needs of Maslow's framework such as need for social interaction and belongingness, need for recognition (reputation) and need for altruism. Moreover, knowledge centric

communities show a strong tendency to cognitive needs, which Maslow added to his stack in later years. Among the main findings, we observed that users differ in their association with various need patterns. Users with high reputation are more likely to stay longer with a community than users dominated by cognitive needs. However reputation seems to have less impact on content quantity as evident from both correlation and factor analysis. Socially motivated users create more content, and engage with other users more frequently compared to users motivated by other needs. Although the relation between need for self-actualisation and community participation is complex to model, we found a strong desire to help other members. Finally, needs are not found to be sequential or hierarchical as proposed in the theory, rather they seem to co-exist in different degrees and intensities at different point of time.

*Limitations:* Limitation of this study is two-fold (1) it entirely depends on system data hence validating the behaviour to need mapping is tricky and may vary with more data available, and (2) the domain of application is restricted to Q&A communities where participation is more controlled and hence many original motivations of participation may not be well reflected, like motivation to get attention or self-promotion or entertainment. In future work we plan to apply our analysis to other type of communities.

## References

- (1) Brewer, D.(2000) Forgetting in the recall-based elicitation of personal and social networks. *Social Networks* vol 22, PP, 29–43.
- (2) Cattell R.B.(1966) The scree test for the number of factors. *Multivariate Behavioral Research*.1(2) pp 245-76
- (3) Chiu, C.-M., Hsu, M.-H., & Wang, E. T. G. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems*, 42(3), pp 1872–1888.
- (4) Crowne, D. and Marlowe, C.(1960) A new scale of social desirability independent of psychopathology. *J. Consulting Psychology* vol 24, pp 349–354.
- (5) Henson RK, Roberts J.K.(2006) Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement*. Vol 66(3)
- (6) Hars, A., & Ou, S. (2002) Working for free? Motivations for participating in open-source projects. *International Journal of Electronic Commerce*, vol 6(3), pp 23–37.
- (7) Joinson, A. N.(2008) Looking at, looking up or keeping up with people?: motives and use of facebook. In *Proc. CHI*. ACM Press, pp 1027-1036. 34.
- (8) Kaiser H.F.(1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement*. Vol 20 pp 141-51.
- (9) K. K. Nam, M. S. Ackerman, and L. A. Adamic. Questions in, knowledge iN? A study of Naver's question answering community.(2009) In *Proc. CHI* pp 779–788. ACM Press
- (10) Krasnova, H., Hildebrand, T., Guenther, O., Kovrigin, A; and Nowobilska, A,(2008) "Why Participate in an Online Social Network? An Empirical Analysis". *ECIS 2008 Proceedings*.
- (11) Lakhani, K. R., & Wolf, R. G. (2005). Why Hackers Do What They Do : Understanding Motivation and Effort in Free / Open Source Software Projects 1 By, 1–27.
- (12) Lampe, C., Wash, R., Velasquez, A., & Ozkaya, E. (2010). Motivations to participate in online communities. *Proceedings of the 28th international conference on Human factors in computing systems – CHI*.
- (13) Maslow, A.H. (1943). A theory of human Motivation. *Psychological review*, vol 50(4), pp 370-96.
- (14) Nam, K. K., Arbor, A., & Ackerman, M. S. (2009). Questions in, Knowledge iN? A Study of Naver ' s Question Answering Community University of Michigan.
- (15) Oded N. (2007) What motivates Wikipedians?. *Commun. ACM* vol 50, nu 11, pp 60-64.
- (16) Oreg, S., & Nov, O. (2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior*, vol 24(5), pp 2055–2073.
- (17) Pett M.A, Lackey N.R, Sullivan J.J.(2003) Making Sense of Factor Analysis: The use of factor analysis for instrument development in health care research. California: Sage Publications Inc.
- (18) Stoeckl, R., Rohrmeier, P., & Hess, T. (2007). Motivations to produce User Generated Content : differences between webloggers and videobloggers pp 398–413.
- (19) Thompson B, Daniel LG.(1996) Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*. Vol 56(2), pp 197-208.
- (20) Thompson B.(2004) Exploratory and confirmatory factor analysis: understanding concepts and applications. Washington, DC: American Psychological Association.
- (21) U. M. Dholakia, R. P. Bagozzi, and L. K. Pearo.(2004) A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing*, vol 21, pp 241–263.